

Is a User-Friendly Diary More Effective? Findings from a Field Test

ERIC FIGUEROA
NHIEEN TO
LUCILLA TAN
JEANETTE DAVIS
SALLY REYES-MORALES

Eric Figueroa, Nhien To, and Lucilla Tan are economists in the Branch of Research and Program Development, Division of Consumer Expenditure Surveys, Bureau of Labor Statistics.

Jeanette Davis is a senior economist in the Branch of Research and Program Development, Division of Consumer Expenditure Surveys, Bureau of Labor Statistics.

Sally Reyes-Morales is a mathematical statistician in the Division of Price Statistical Methods, Branch of Consumer Expenditure Surveys, Bureau of Labor Statistics.

Diary surveys are often used to collect information on daily activities such as consumer spending. They are particularly useful for collecting daily records of small frequently purchased items, which are normally difficult to recall.¹ The Consumer Expenditure (CE) survey, sponsored by the U.S. Bureau of Labor Statistics (BLS), with data collected by the U.S. Census Bureau, uses a diary survey to collect data on weekly household expenditures.

Recent efforts to improve the performance of the CE diary survey have focused on designing a more user-friendly form. Such a form would have a simpler recording scheme and be more attractive in appearance than the form currently used in production. Several prototype diaries were developed and refined with the use of feedback from survey respondents, field interviewers, and program staff.² On the basis of this feedback, CE management selected one of the designs (the Redesigned Diary) for field testing. This diary was intended to stem declining response rates and improve data quality by reducing respondent's burden associated with

¹ S. Sudman and N. Bradburn, *Asking Questions*, (San Francisco, Jossey Bass Publishers, 1982).

² J. Davis, L. Stinson, and N. To, "Creating a 'User-Friendly' Expenditure Diary," *Consumer Expenditure Survey Anthology* (Bureau of Labor Statistics, 2003), Report 967, p. 3.

the diary now used: the Production Diary. The Redesigned Diary is smaller and shorter than the Production Diary, has a simpler organization, and highlights important instructions and examples.

The Redesigned Diary was tested in the field from October through December of 2002.³ The primary objective of this field test was to compare the response rates and data quality obtained from the Redesigned Diary with those obtained from the Production Diary. The results showed no statistically significant difference between diary forms in completion response rates and only a few significant differences in expenditure means and allocation rates. (The latter measure the proportion of expenditures requiring further processing because they are reported with insufficient detail.⁴)

However, the Redesigned Diary performed statistically significantly better than the Production Diary in a majority of tests pertaining to the collection of item attribute information needed for

³ A field test is designed to reproduce data collection conditions as closely as possible to those in the production environment.

⁴ Allocation is an adjustment performed on expenditure entries that do not identify individual items at the required level of detail (for example, a report that says "groceries \$150," rather than listing the specific items purchased and the price of each). This type of entry requires additional processing to assign the aggregate expenditure to target items.

classification.⁵ In addition, the Census Bureau field representatives who worked on the field test expressed a strong preference for the Redesigned Diary because of its more attractive layout and simpler recording scheme.

On the basis of the field test results, it was decided to continue research on the Redesigned Diary before implementing it in production. The focus of the research was to test modifications to the Redesigned Diary that would increase reporting of expenditure levels for *food away from home* and reporting detail for *food for home consumption*.

Background

Diary Survey Instruments. Two paper-and-pencil questionnaires are currently used to collect diary data. The first is the Record of Daily Expenses, the actual diary form. This is a self-reporting form on which respondents record a detailed description of all expenses for their consumer units (CUs) for two consecutive 1-week periods. (Data collected each week are considered independently.) The diary is divided by day of purchase and by broad classifications of goods and services—a breakdown designed to aid the respondent in recording daily purchases. Currently, the major classifications are as follows:

- Food away from home
- Food for home consumption⁶
- Clothing, shoes, and jewelry
- All other purchases and expenses

Each classification is further divided into numerous subcategories within which the items reported are subsequently coded by the Census Bureau. Thus, BLS can aggregate individual purchases for representation in

⁵ *Attribute information* is needed to classify items; the percentage of entries missing such information measures the portion of entries for which respondents did not provide the needed attribute information (for example, a respondent who reports “peas,” but does not provide attribute information on the type of package—fresh, frozen, or canned).

⁶ Includes food and beverages purchased as gifts.

the Consumer Price Index and for presentation in statistical tables.

The second questionnaire used to collect diary data is the Household Characteristics Questionnaire, used to record information pertaining to age, sex, race, marital status, and family composition, as well as information on the work experience and earnings of each member of the consumer unit. This socioeconomic information is used by BLS to classify the CU for the publication of statistical tables and for economic analysis. Since 2003, the Household Characteristics Questionnaire has been administered with the use of computer-assisted personal interviews (CAPIs).

Redesigning the Diary Form. The objective of redesigning the diary was to produce a more user-friendly form to encourage higher response rates and more accurate reporting. BLS and the Census Bureau began developing the Redesigned Diary in 2000. Findings from focus groups were used to define the features of a user-friendly form: a form that is easier to understand, less complicated to navigate, simpler to complete, and looks more attractive than the Production Diary. Through a series of cognitive tests of several prototype diaries designed with these user-friendly features, one—the Redesigned Diary—was selected for testing in the field.

Following is a summary of the differences in the features of the Production Diary and the Redesigned Diary:

- *Smaller physical size.* The Redesigned Diary is smaller (8 ½” × 11”), has fewer pages (44), and is in portrait format. In contrast, the Production Diary is 14” × 8” with 66 pages and is in landscape format.
- *Simplified layout.* The Redesigned Diary has a simpler organization than the Production Diary. In the Production Diary, each day’s reporting space consists of seven pages, broken down into broad classifications and numerous

subcategories. In the Redesigned Diary, each day’s reporting space is reduced to four pages, also broken down into broad classifications, but without subcategories, simplifying the respondent’s task and the form’s appearance.

- *Clearer instructions and examples.* The Redesigned Diary’s instructions are formatted so topics are easier to find:

1. The Production Diary’s instructions are evenly spread over two pages, divided into eight topics, distinguished by their titles, which compete with numerous subtitles. The Redesigned Diary’s instructions are also contained on two pages, but the different topics are more easily distinguished from one another. The information is grouped into three topics, graphically set apart from one another through the use of frames and by means of title blocks in large fonts.
 2. A section titled “Frequently Asked Questions” was added to the Redesigned Diary. This section answers common questions asked about the diary-keeping task and is found on an easily accessible flap on the diary’s back cover. Examples of expenditures are contained on a flap on the front cover. Both flaps can be used as bookmarks to help the respondents keep their place.
 3. Compared with the Production Diary, the Redesigned Diary has a greater variety of examples, focuses on difficult cases, and highlights important data entry instructions and examples by using color, white space, boldface text, and superimposed balloons.
- *More check boxes to facilitate the recording task.* In contrast to the Production Diary, the Redesigned Diary has more check boxes, allowing respondents to classify expenditures more easily.

- *A more current and appealing look that still maintains a professional and official quality.* The Redesigned Diary uses color and photos to cue respondents and to make the diary more appealing. The Production Diary is printed in black and green on white paper and has no photos.

The 2002 Field Test

Sample Design. To assess the performance of the Redesigned Diary, a field test was conducted from September through December 2002. In addition to the redesigned form, a CAPI version of the Household Characteristics Questionnaire was tested. This alternative replaced the paper-and-pencil version of the questionnaire formerly used in production.⁷

The field test design included both test (Redesigned Diary) and control (Production Diary) samples. Both samples used the CAPI Household Characteristics Questionnaire. To create the samples, the Census Bureau selected 1,800 households from a previously unused supplemental sample. These sample units were drawn from 9 of the 12 Census regions.⁸ The test sample of 1,200 households received the Redesigned Diary, and the control sample of 600 households received the Production Diary.

As the field test proceeded, significant demographic differences were found between the test and control samples. The largest such differences identified were in the proportions of owners and renters. In the test sample, these proportions were close to those found in the general population. In the control sample, the proportion of renters was higher than that found in the general population. In addition, renters in the control sample had significantly lower incomes than renters in

the test sample. Because these characteristics affect expenditure levels, the disparities weakened the control sample's usefulness for comparisons with the test sample output.

In anticipation that the control sample would not be large enough to provide meaningful estimates, a production sample was selected for comparison with the test sample. The production sample was drawn from concurrent production data restricted to the regions, Metropolitan Statistical Areas, and sample frames used to draw the field test sample. The resulting sample consisted of 2,703 households.

Given the aforementioned differences in the demographics between the test and control samples, the authors chose to focus on comparisons between the test and production samples. Although the production data had been collected without the CAPI component, the demographic consistency of its data with the test sample was thought to make it a better subject for comparison.

Measures of Effectiveness. Our research goal was to compare the effectiveness of the Redesigned Diary with that of the Production Diary. Our null hypothesis states that they are equally effective. Our alternative hypothesis asserts that one diary is more effective than the other.

The more effective diary must have the following two attributes:

1. *Higher completion response rates.* Completion response rates measure the percentage of all eligible diaries successfully placed and completed.⁹
2. *Higher mean dollar expenditures per CU in the two food expenditure categories: food away from home and food for home consumption.*¹⁰

These two criteria were selected, respectively, because of concern over the declining response rates in the CE survey and the importance of the diary as the major source for data on food expenditures. It would also be desirable if a diary produced higher mean expenditures in the two nonfood expenditure categories, produced relative expenditure shares¹¹ consistent with the pattern in current production data, and had lower percentages of entries missing attribute information. However, it is sufficient for one diary to be judged more effective than the other if it meets the foregoing two criteria.

In addition to the quantitative analyses on the field test data, two other analyses were undertaken to evaluate the diary:

1. *A content analysis of the Redesigned and Production Diaries.* The objective of a content analysis is to compare the overall quality of entries in the diaries: Whether entries were recorded properly and clearly and whether relevant check boxes were marked. Ten percent of diaries were randomly selected for content analysis, ensuring coverage in the three areas: Single and multiperson CUs, diaries from Weeks 1 and 2, and diaries from all geographic regions.¹² A total of 47 Control Diaries and 81 Redesigned Diaries from the months of September and October were reviewed.
2. *A debriefing of field representatives.* Field representatives who participated in the field test were given an opportunity to share their impressions and reactions. In December 2002, a debriefing questionnaire was sent to those who participated in the field test. The response rate for this question-

⁷ After further refinement, the CAPI version was introduced into production in 2003.

⁸ The nine Census regional offices that participated in the field test were Atlanta, Boston, Charlotte, Chicago, Dallas, Denver, Detroit, Philadelphia, and Seattle; excluded were New York, Los Angeles, and Kansas City.

⁹ *Eligible* housing units are those in the designated sample, less housing vacancies, housing units under construction, housing units with temporary residents, destroyed or abandoned housing, and units converted to non-residential use.

¹⁰ The latter category includes food and beverages purchased as gifts.

¹¹ The relative share of each of the four expenditure classifications is the percentage of total expenditures that each constitutes.

¹² The geographic regions are the Northeast, Midwest, South, and West.

naire was 86 percent. A total of 17 field representatives representing the 9 Census regional offices participated in a 1-day debriefing in January 2003.

Determining Significant Differences. Statistical tests were performed to measure significant differences in the output of the Redesigned and the Production Diary. For the Redesigned Diary field test, variances were calculated using the method of “random groups.”

To obtain the random groups required for statistical analyses of the test and production samples, the CU universe was randomly divided into 10 groups called replicates, with each replicate containing approximately 10 percent of the universe. Each statistic of interest (such as mean expenditure, response rate, and relative importance) was computed separately for each replicate, as well as for the full sample.

Then the variance for the statistic is estimated by

$$\text{Var}(\bar{x}) = \frac{\sum_{r=1}^{10} (\bar{x}_r - \bar{x})^2}{10(10-1)} 5,$$

where

\bar{x} = the full sample statistic of interest

and

\bar{x}_r = the statistic for the r th replicate.

The standard error is estimated by

$$\text{SE}(\bar{x}) = \sqrt{\text{Var}(\bar{x})}.$$

To determine whether the statistic of interest was significantly different between the test \bar{x}_{Test} and production ($\bar{x}_{\text{Production}}$) samples, z -scores (Z) that allow a statement of statistical significance were calculated with the formula

$$Z = \frac{|\bar{x}_{\text{Test}} - \bar{x}_{\text{Production}}|}{\sqrt{\text{Var}(\bar{x}_{\text{Test}}) + \text{Var}(\bar{x}_{\text{Production}})}},$$

where $\text{Var}(\bar{x}_{\text{Test}})$ and $\text{Var}(\bar{x}_{\text{Production}})$ are the variance of the test and production statistics, respectively.

If $|Z| > 2$, then the difference between the statistics of interest is statistically significant.

Findings

On the basis of comparisons between the test and production samples, the data yielded the following results:

Response rates. No significant difference in the response rates for completed diaries was found. (See table 1.) Compared with the refusal rate in the Redesigned Diary, the refusal rate in the Production Diary was significantly higher. However, the Redesigned Diary also had a significantly higher rate of incomplete interviews for “other” reasons, perhaps due to the more stringent placement dates enforced by CAPI.

Expenditure means. In the Redesigned Diary, expenditures were significantly lower for *Food Away from Home*, but significantly higher for *Clothing, Shoes, and Jewelry*. In terms of expenditure shares—the percentage of total expenditures spent on each component—only *food away from home* was significantly lower in the Redesigned Diary. These results may be due to new titles¹³ in the Redesigned Diary for *food away from home and food for home consumption*. Because of the difference in titles, respondents using the Redesigned Diary may have thought they should omit from the *food away from home* section some expenditures that respondents using the Production Diary thought should be included.

Allocation rates. In the Redesigned Diary, the percentage of expenditures for *Food Away from Home* coming from allocation was significantly lower than that in the Production Diary. The dif-

¹³ In the Redesigned Diary, the *food away from home* and *food for home consumption* sections were retitled, respectively, “Food & Drinks from Food Service Places” and “Food & Drinks from Grocery and Other Stores.”

ference may be largely a reflection of the effectiveness of the additional check boxes in the Redesigned Diary. No other significant differences were found.

Percentage of missing attributes. Three of the five tests (*meal type, alcohol type, and gender*) showed significantly lower rates of missing attributes in the Redesigned Diary compared with the Production Diary. As with *food away from home*, this phenomenon may be due largely to the effectiveness of additional check boxes. One test (*package type*) showed significantly lower results in the Production Diary, and one (*age*) showed no difference between the diaries.

Content analyses. On the basis of the diaries that were manually reviewed, it was not apparent that one type of diary had consistently higher error rates than the other. (See table 2.)

Debriefings of field representatives.

- Survey of Census Bureau field representatives who administered the field test. The field representatives expressed overwhelming support for the Redesigned Diary. When asked to compare the two diaries on several criteria (overall impression, ease of administration, ease of respondent use, layout design, complete interviews obtained, accurate data obtained), a majority of the field representatives consistently gave the Redesigned Diary favorable ratings and gave the Production Diary neutral or negative ratings.
- In-person debriefing of 17 representatives. The majority of the field representatives thought that the format of the Redesigned Diary, with fewer categories, effectively reduced respondent burden. They believed that respondents were more likely both to record in the diary and to persevere with recording entries through the second week.

Conclusion

The findings of the diary field test did not allow us to reject the null hypothesis. Thus, both the Redesigned Diary and the Production Diary are equally effective. No significant difference was found in the test of completion response rates. Results were mixed for tests of mean expenditures in the two food categories: the Redesigned Diary had significantly lower expenditures than the Production Diary had for *food away from home*, and there was no significant difference between the diaries in *food for home consumption*. Higher results on both tests were necessary for either diary to be judged more effective than the other.

The Redesigned Diary performed significantly better in a majority of tests

having to do with missing attribute information. Taking into account all test differences—whether significant or not—we find that the Redesigned Diary produced higher expenditure means and lower allocation rates in three of the four expenditure categories. In addition, the field representatives who worked on the field test expressed a strong preference for the Redesigned Diary.

Further Research

The Redesigned Diary's weak areas merit additional research. The expenditure means in the *food away from home* section were lower in the Redesigned Diary than in the Production Diary. Cognitive work is needed to determine whether the titles used in each diary

are confusing to respondents, possibly leading to incorrect items being entered.

Additional research also is needed to develop effective cues to encourage more detailed reporting in the *food for home consumption*, the *clothing, shoes, and jewelry*, and the *all other purchases and expenses* sections. The cues should not be overwhelming or add significant amounts of respondent burden.

The authors would like to acknowledge the following BLS employees who contributed to this analysis: Jeff Blaha, Richard Dietz, Tammy Hagemeyer, William Mockovak, Troy Olson, Mary Lynn Schmidt, Linda Stinson, David Swanson, Clyde Tucker, and Wolf Weber. ■

Table 1. Comparison of data from the Redesigned and Production Diaries

Characteristic	Test (CAPI and Redesigned Diary)	Production (Production Diary)	Significant difference
Response rates (percent):			
Completed	74.5	75.2	—
Eligible CUs who did not complete interview because—			
refused	11.9	17.9	****
not home	5.0	4.3	—
other	8.6	2.6	****
Mean expenditures (dollars):			
All expenditure categories	371	359	—
Food for home consumption	64	64	—
Food away from home	37	41	**
Clothing, shoes, and jewelry	39	33	**
All other purchases and expenses	231	221	**
Allocation rates(percent of expenditures from allocated items):			
All expenditure categories	17.6	20.8	—
Food for home consumption	24.3	26.3	—
Food away from home	18.3	49.5	****
Clothing, shoes, and jewelry	22.2	17.5	—
All other purchases and expenses	15.6	16.2	—
Missing attributes (percent of entries missing attribute information):			
Package type.....	7.2	4.7	**
Meal type	2.8	30.3	****
Alcohol type	9.8	16.6	**
Age	17.7	21.4	—
Gender	16.4	21.4	**

NOTES: Statistical significance based on Z-score: ** $2 \leq \text{abs}(Z) < 3$, *** $3 \leq \text{abs}(Z) < 4$,**** $\text{abs}(Z) \geq 4$. Dash indicates no significant statistical difference.

SOURCE: The Consumer Expenditure Survey Redesigned Diary field test, September–December 2002.

Table 2. **Content analysis of the Redesigned and Production Diaries**

Characteristic	Redesigned Diary (in percent)	Production Diary (in percent)
Error rate of illegible entries (cannot read, due to handwriting):		
Food away from home	0.0	0.0
Food for home consumption4	.2
Clothing, shoes, and jewelry0	.0
All other purchases and expenses2	.0
Error rate of unintelligible entries (can read, but cannot tell what the entry means):		
Food away from home6	.0
Food for home consumption9	5.5
Clothing, shoes, and jewelry0	.0
All other purchases and expenses9	1.8
Error rate of missing description fields:		
Food away from home7	.0
Food for home consumption0	.0
Clothing, shoes, and jewelry0	.0
All other purchases and expenses0	.0
Error rate of missing total-cost fields:		
Food away from home0	.0
Food for home consumption0	.2
Clothing, shoes, and jewelry0	.0
All other purchases and expenses0	.6
Error rate of missing alcohol check marks (when alcohol is described or cost is given):		
Food away from home0	3.4